DOCUMENT RESUME

ED 137 413 TM 006 222

AUTHOR Hayman, John; Rayder, Nick
TITLE Evaluation and Public Policy.
PUB DATE [Apr 77]
NOTE 26p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Conferences; Decision Making; *Evaluation; *Evaluation Methods; *Evaluation Needs; Feedback; *Humanization; Information Needs; Problems; Program Evaluation; *Public Policy; Racism; Sex Discrimination; Social Values; Systems Concepts

ABSTRACT

Educational evaluation has been generally ineffective in improving educational practice, and many evaluations have dehumanizing aspects. A conference was held at Berkeley in April, 1976, to consider this situation. Participants included professional evaluators, teachers, parents, students, and interested others. One outcome was this hypothesis: The efficacy of information produced in an evaluation effort varies inversely with the number of organizational levels that the action the information describes is removed from the decision process the information is intended to influence. In support, it is noted that any information has a syntactic aspect, a semantic aspect, and an effectiveness or influence aspect. It is then argued that the syntactic aspect is violated in trying to detect small effects by aggregating data across organization levels; that greater error variance is produced in such an effort; that the most useful information semantically is that from the same organizational level as the user; and that information from other levels is certain to be less motivating. Recommendations include humanizing evaluation, involving the evaluated, removing racism and sexism from evaluation efforts, use of more appropriate evaluations by teachers, increasing the relevance of evaluation to decision makers, and organizing a public constituency for change of evaluation philosophy and practice. (Author/MV)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED137413

TM006 222

EVALUATION AND PUBLIC POLICY

By

John Hayman
Alabama Information and Development System

Nick Rayder
Far West Laboratory for Educational Research and Development

I. The Issue

Criticism of educational evaluation has been increasing the past few years. One reason is the general ineffectiveness of evaluation in improving educational practice, and a second, perhaps more damming one, is the dehumanizing aspects of many of the evaluations which have been conducted. Small groups of concerned evaluators, such as the North Dakota Study Group, have attempted for several years to focus attention on critical issues such as the testing of black children, and the inappropriate selection of children for special education programs. Generally, the concerns of these groups have failed to dent the formal agendas of orthodox forums.

While most professional evaluators have debated their own esoteric concerns in isolation from those affected by their work, however, a growing number of the parents, students, teachers and school administrators have come to perceive that educational evaluation, as it is currently practiced in many situations, may cause problems instead of helping to solve them.

Paper presented at the Annual Convention of the American Educational Research Association, New York City, April, 1977.

The situation calls for a re-examination. Stake (undated) comments that:

> Maybe the whole idea of evaluation should go in the circular file. School districts in Ohio, Illinois, and everywhere else have problems. Many of the problems are aggravated by people who come to evaluate. We do not know whether or not evaluation is going to contribute more to the problem of education or more to the solutions.

This sort of concern occurs at the same time that there is increasing demand in the society for accountability, for demonstration by educators that they are in fact producing the outcomes they are supposed to--however vaguely defined these outcomes may be. Being more accountable means producing hard data on the "bottom line" to most people, and thus there appears to be a dilemna between the need for more and better information on the one hand and the consequences of collecting it by current methods on the other.

All of this is a precursor of more difficult dilemnas to come. Society is moving rapidly into the so-called "Age of Information," which is characterized by expanding complexity in all aspects of life, closer interrelatedness among groups and social systems throughout the world, and enormous increases in the knowledge base. The demand for more systematic approaches to education and for greater use of technology and of information on individuals in the educative process will grow. Movement in this direction makes even more imperative the concern of humanists that the basic values of society be preserved. Maintaining rights of privacy in the face of growing technology, for example, is a fundamental and extraordinarily difficult task.

Riles (1973) notes that the problem in a sense is one of balance:

> Our society and our schools need a better balance of values. That means the technicians, the scientists, the engineers must concern

> themselves more with the human consequences of their research and achievements. And it means that the humanists, the artists, the educators must concern themselves more with the practical, realistic and technical planning necessary to achieve their purposes.

It was against this background that the conference reported in this paper was planned and conducted. The purpose in general was to develop a clearer understanding of the goals of evaluation in relation to the values of society and hopefully to create the impetus for integrating positive change into the evaluation process in education.

II. Planning and Developing the Conference and Its Purposes

The Conference's development can be traced to mid-1975, when individuals and groups involved in evaluation in the San Francisco Bay Area began meeting with each other, and discovered that they shared common concerns. These meetings promoted an awareness of the existence of a diversified national constituency which was critical of the current state of educational evaluation and which was either already articulating or was ready to articulate humanistic views within the field.

The initiators decided to attempt to convene a significant portion of this constituency in an atmosphere conducive to examination of the issues and to dialogue about the future of educational evaluation. Interest in current evaluation practices was intense enough to forge an ad hoc coalition of more than a dozen members who worked without remuneration for more than a year to develop a conference which would attract people from various perspectives, professional and life styles, and which would represent a meaningful alternative to the traditional context in which evaluation has been discussed.

Many Bay Area evaluators had been feeling frustrated about the de-humanization of evaluation, as expressed in the evaluation of national Title I

and Follow Through Programs and of public education in general. At various conferences, when they tried to voice these frustrations, they were confronted with the absence of a professional ethic which could accommodate humanistic evaluation issues.

The absence of a professional ethic did not mean absence of interest in the topic within the profession, however. At social hours and other informal gatherings, numerous colleagues were found who were troubled by such dehumanizing evaluative trends as discounting all educational outcomes except those easily quantified through standardized testing.

In the fall of 1975, Paul Chaffee, of the California Council for the Humanities in Public Policy, was actively meeting leading education experts in the Bay Area to discern the policy issues which were currently being debated. The Council, a state-based agency of the National Endowment for the Humanities, funds public programming of policy issues and asks grantees to include academic humanists in the dialogue. Chaffee indicated that the Council would happily receive an application to fund a forum in which the major issues relating to educational evaluation could be raised.

Chaffee was in contact with Nick Rayder, an evaluation specialist with Far West Laboratory in San Francisco. Together with Bill Baker of the Alameda County School District, Rayder began communicating with the Bay Area Consortium, composed of school district representatives from six Bay Area counties and programs such as Berkeley Experimental Schools Project, Far West Laboratory and University of California at Berkeley Field Services.

In the summer of 1975, Consortium members envisioned a conference to follow the 1976 American Educational Research Association conference in San Francisco. Members felt that AERA often bypassed the real issues of

educational evaluation. A grant application for the follow-up conference was submitted to the California Council for the Humanities and was successful.

The Berkeley Experimental Schools Project agreed to act as a co-sponsor of the conference, providing substantial financial and in-kind support. A variety of groups, including the school districts of San Francisco, Richmond and Alameda County, the University of California Field Services, The Institute for Scientific Analysis, Far West Laboratory, and the San Francisco Classroom Teachers Association provided in-kind services. At a later stage, Phi Delta Kappa supported the dissemination of project material.

To constructively attack the problems within a conference framework, the planners decided on the following organizational principles:

1. The conference should include concerned professionals from a range of advocacy positions in order to establish a rigorous context in which to consider some emotionally-charged issues and to convey to professional evaluators that the movement for evaluation reform is widespread (if not yet clearly focused).
2. Concerned individuals from the full variety of evaluation-impacted life and career roles should be included, and the atmosphere of the conference should be informal enough to permit each individual to feel qualified as an "expert" about his or her own evaluation-related experience. The conference should not be dominated by evaluation professionals.
3. The conference's format should reflect an appreciation of pluralistic values, should be flexible and should be experienced-based in order to convey a sense of urgency around issues.
4. The conference participants should feel that resources and expertise

sufficient to make a major impact on evaluation procedures were present at the conference, and that the question to be faced was, "What do we do?" rather than "What can we make others do?"

The conference was held as scheduled in April 1976. It represented the first major attempt to engage evaluation analyzers in dialogue with each other, with evaluation practitioners, with humanist scholars, and with funding sources.

III. Conference Proceedings

Several large group experiences helped set the tone of the conference and provided a basis for broad scale interactions among participants.

Immediately after registration on the opening day, the 150 conference participants gathered to view a frightening slide-tape presentation concerning the year 2000. The narrator described a United States on the brink of total disintegration. The only hope, the viewers learned, lay in rigorous selection of the cream of the nation's youth for special school and advancement to leadership positions.

Each participant was given a packet consisting of information about 10 hypothetical youthful candidates for special educational treatment in the year 2000. Just two of the 10 candidates could be admitted to the special program, and participants were asked to make a choice. The selection, it was stated, would determine not just the educational future of individual candidates but also the general well being of the nation and the allocation of considerable funding. Data on candidates included demographic information, physical descriptions, intelligence quotients, social quotients, health evaluations, and information about goals, interests, skills, and political beliefs.

Participants worked in small groups, which made choices. Then the slide show recommenced and it was revealed that all of the applications had psuedonymously described gifted individuals. Most of the groups had excluded Albert Einstein (IQ of 82, very poor school adjustment, etc.), Isadora Duncan (low social quotient, behavior problem), Eleanor Roosevelt (erratic, withdrawn, seeks attention, fails often) and four others of the most illustrious individuals of our times.

The point was well made that judgmental evaluations -- evaluation which separates good students and programs from bad -- is dangerous. The consequences of misjudging or harmfully catergorizing human beings can be disasterous. It is clear that the issue is complex and that there are no quick answers.

A second large-group experience was a presentation by the Majumbe Dancers of black rhythms from across the centuries. This group, through their performance, illustrated that American educators and evaluators must recognize that the dominant culture's definition of valid learning experiences and styles is not all-encompassing or, indeed, sufficient.

The third major large-group experience, a Saturday panel discussion on the topic, "Does Educational Evaluation Influence Public Policy? To What Extent? At Whose Expense?" forced awareness of differing priorities for change. By the time the discussion was over, it was no longer possible to entertain the illusion that opposition to standardized testing is a sufficient common platform for reform of educational evaluation. Dispelling that illusion and recognizing the diversity of viewpoints and interests that can be brought to bear upon evaluation by those <u>outside</u> of the evaluation "establishment" seem a necessary starting point for improvement.

Other experiences involving conference participants as a group included Chinese, Mexican, and soul-food meals, a reading of poetry about children, social hours and several summary sessions at which the progress of the conference was assessed. The conference schedule and format were adapted to meet the evolving conditions. As the participants faced issues from a variety of perspectives, personal interaction grew and abstraction decreased. This process was intensified when participants entered small-group discussions, which were the other major activity of the conference.

Twelve workgroups with about a dozen participants each met four times during the conference, with each session lasting about two hours. A conscious effort was made to provide for diversity of experience within each group. A facilitator from the conference planning committee provided initial impetus and structure as the sessions got underway.

The basic conference impetus toward reform of evaluation practice did not preclude a variety of separate agendas for such reform. For some, the workgroups were a frustrating experience, which dashed hopes of easy alliances toward "obvious" objectives. For others, the workgroups provided a vital advance toward coping with the complexity of the forces involved in the evaluation controversy.

In the conference's final session, the individual workgroups reported to the whole. Virtually every group achieved concensus for its recommendations, and it was clear that the energy for reform had maintained its force despite differing priorities. A widespread recognition had emerged that there are many problem areas, many routes to reform.

Building on the workgroup process, a number of conference participants called an informal meeting one evening in an attempt to determine if a conference-wide concensus could be forged around any course of reform action. More than 50 participants met late into the night, but at that meeting, as throughout the conference, those favoring a broadbased thrust toward short-term goals found themselves at odds with those emphasizing long term, systemic policy change.

Those seeking to reform particular testing instruments and those opposed to all standardized testing could not reconcile their differing priorities. Eventually, the participants agreed that the only feasible course was to identify problem areas within educational evaluation and provide a framework

for conference participants, and other interested persons, to pursue their various change agendas.

The point should be made that the conference was not anti-evaluation -- just anti educational evaluation which is useless or harmful or both. Indeed, there was recognition that evaluation which is positive, which assists learners and educators to meet goals, and which is humane in its methods and outlook is essential. The facts that it is essential and that such a large proportion of educational evaluation to date has failed to meet these criteria make more crucial the critical considerations which emerged. If evaluation were not so important, then the issue would be simple; do away with it, except perhaps as a scholastic exercise. This cannot and will not be the outcome, however, so the only meaningful course is to press to make evaluation in education more effective.

IV. A Technical Outcome: Selecting Functional Analysis Units

A. The Cross-Levels Hypothesis

In relation to increasing the effectiveness of educational evaluation, a technical point emerged during the conference which seemed quite simple and logical at the time, but which, on further reflection of the authors, seems to be more fundamental that it first appeared. It might be stated in hypothesis form as follows:

> The efficacy of information produced in an evaluation effort varies inversely with the number of organizational levels that the action the information describes is removed from the decision process the information is intended to influence.

Evaluation is a process which produces information, as we all know, and the information is intended to impact decisions. Any evaluative information is produced by observations of the behavior of some entity, some unit of analysis.

What the hypothesis says is that information resulting from evaluation will be more useful, in the sense of impacting decisions, the closer the decision maker is, organizationally, to the analysis unit. Information describing student performance in a classroom, according to this hypothesis, will be most useful to the classroom teacher, less useful to the principal, still less useful to the district superintendent, etc.

Looking at it from the other direction, the hypothesis says that, if you are a decision maker at some level in education and you need evaluative feedback to help you perform your job and meet your goals, the most useful feedback will be in terms of analysis units at your organizational level or at the levels immediately adjacent to you. If you are the US Office of Education, for example, and you are attempting to evaluate the results of some program -- say Title I --

then the most useful information will come from descriptions of organizational behavior at state government level.

The hypothesis, considered in these terms, turns out to be very interesting, for we are all aware of general failure to produce any definitive evaluations of Title I at the national level -- in spite of millions of dollars spent on the effort over some 12 years. In fact, a good many of the national evaluations of federally-funded programs would have to be classified as failures in terms of producing information helpful to decision making.

Why has there been failure, when some of the best brains around have given their best to the effort? Is it because evaluation is a field still in its adolescense? Is it because the field is so fragmented and so much controversy exists within it? All of these relate to the problem, no doubt, but our hypothesis suggests a more basic reason. Most of the national evaluations have relied heavily on securing data on students at the classroom level, and then aggregating it up through various levels until it is finally put together in one large set nationally. If the hypothesis is valid, then one would predict in advance that such an evaluation effort would be of little value!

B. Rationale

The hypothesis emerged intuitively at the conference. After thinking through its consequences and realizing that it could be an important principle, the authors did some background work to see if a theoretical rational could be built to support it. What follows in this section, therefore, was not produced at the conference, but is a post hoc attempt to demonstrate why the cross-levels hypothesis may be valid. We will be necessarily brief here, with the intention of developing the argument more fully in the future.

1. Information and Its Purposes

We first need to define "information" more carefully. According to Weaver (1963), information can be conceived in three aspects -- according to its purely technical or syntactic qualities, according to its semantic qualities, or according to its effectiveness in influencing action.

The technical or syntactic aspect deals with the engineering matters of transmission and receiving symbols, and the major concern is the accuracy with which the transmission can be accomplished. In the semantic aspect, the concern is with the understanding by the receiver of the meaning of the symbols which have been transmitted. The effectiveness or influence aspect involves the capability of a message or set of data to motivate human action.

These operate in a hierarchy; accurate transmission is essential to understanding, and understanding in turn is essential to effectiveness. Note, however, that planning for the use of information to support decision processes must start at the top, that is, with the determination of what is needed for influence.

The function of information is to reduce uncertainty related to specific needs. In organizations, information conveys objectives, plans, policies, and procedures to all components (subsystems). In this context, it serves both as a process activator to communicate requirements and as feedback to ensure that the communication has been received, understood, and acted upon (Shrode and Voich, 1974, pp. 420-421). Information also performs another extremely crucial function of linking the organization to its environment. Information, therefore, reduces uncertainty about environmental challenges which will be encountered, about action alternatives for meeting these challenges, about choices which have been made, about the extent to which choices have been implemented and about outcomes.

Information can fail to perform its functions for several reasons. It can be transmitted inaccurately. It can be misunderstood. It can be ignored. It can fail to reduce uncertainty, or it can address issues which are not relevant.

2. Applicable Theory

Given this general background, there are four theoretical propositions which support the cross-levels hypothesis of evaluation efficacy.

Detecting Small Signals Amidst Great Noise.--In their review of the Handbook of Evaluation Research, the Stanford Evaluation Consortium (1976) comments on the difficulty of determining effects in social interventions.

> The relationship between the processes of social intervention and evaluation has proven more complex and disconcerting than imagined. . . . . (The causes of social ills) are usually multiple and only a few of them are affected by any one intervention. When many factors determine outcomes, the evaluator's task is to detect a small signal amidst a great deal of noise. Moreover, when factors combine multiplicatively, a standard intervention may be necessary but insufficient, and there may be no signal to detect (pp. 12 and 13).

Educational evaluators almost invariably work in complex situations and attempt to assess the effects of various types of social interventions. The factors determining outcomes are multiple, and there is noise in the system -- with questionable measuring devices, for example, and with interference from other (unmeasured but relevant) variables. Under the best of conditions, they are unlikely to produce information which can definitively pinpoint cause-and-effect relationships.

There is a high probability, therefore, of failure in the first aspect of the information hierarchy referred to previously, i.e., in the syntactic sense. Incorrect symbols are likely to be transmitted. As we shall see below, this situation is exacerbated when evaluative information must flow across levels.

Greater Variance and Lost Meaning in Cross-Level Generalizations.--The term "unit of analysis" suggests itself when considering the hypothesis. In one sense, what we are saying is that the efficacy of educational evaluations can be increased if evaluators will choose proper units of analysis. This is in turn suggested by certain concepts in General System Theory.

As Miller notes (1975, p. 351):

> The universe contains a hierarchy of systems, each higher level of system being composed of systems of lower levels. Atoms are composed of particles, molecules of atoms . . . . organisms of organs, groups (e.g. herds, flocks, families, teams, tribes) of organisms, organizations of groups, societies of organizations, and supranational systems of societies and organizations.

Miller states that it is important to follow one procedural rule in reference to system level to avoid confusion. "Every discussion should begin with an identification of the level of reference, and the discourse should not change to another level without a specific statement that this is occurring (P. 21)"

Another very important point is that "Cross-level generalizations or hypotheses which may apply to two or more levels will, ordinarily, have greater variance than the other sorts of generalizations, since they include variance among types and among individuals (p. 352)."

Still another key consideration is:

> The more complex systems at higher levels manifest characteristics, more than the sum of characteristics of the units, not observed at lower levels. . . . Significant aspects of living systems at higher levels will be neglected if they are described only in terms and dimensions used for their lower-level subsystems and components (Miller, 1975, p. 353).

If I start trying to consider the behavior of John Doe, teacher, I can describe the relevant variables at any number of levels. I could measure and talk about the biochemical reactions taking place in his cells. I would produce a terribly large quantity of data, of course, and I would experience considerable difficulty discerning from it the way Mr. Doe reacted to a student's request for help. His biochemical processes are relevant; they are just too many levels re-

moved from my level of reference, however, for the semantic sense to be satisfied. I have great difficulty understanding what the data mean in terms relevant to my concern, even if all of the data have been measured and transmitted accurately. And all of it put together will not tell me all about the total behavior of the referent system, Mr. Doe.

Even if I had some great data-reducing mechanism which could integrate it all and perform this degree of cross-level aggregation, I would still have the problem of error variance being introduced at each level. In a situation where the relevant signals are already weak and difficult to detect, introduction of more variance is likely to be fatal to any chances of ever getting the information to perform its influencing function.

Updating a System's View of Itself.--As noted above, a major function of information is to reduce uncertainty about an organization's current state. There are certain "key" variables or attributes, the levels of which relate to success in achieveing goals, and persons in decision roles need to know periodically the status of these attributes so that they can take actions which will be goal-directed.

The term "cybernetic" has been applied to this use of information, and the model is somewhat as follows:

1. The system has recognized goals.
2. Characteristics which relate to the achievement of these goals are understood, at least to some extent.
3. These characteristics can be measured validly and transmitted accurately and meaningfully.
4. Different decision makers are given information on characteristics relevant to their roles and levels of interest.
5. Based on the information, they make decisions and take goal-oriented

actions.

6. New information on the characteristics is then fed back, the results of the previous decision can be assessed, and further actions based on the new information can be taken. In this manner, the system is guided to its goals.

This model applies at all levels of self-organizing systems, and a key point is that the information must be relevant to the particular decision maker's concerns. Suppose, for example, that I am a classroom teacher, and I am concerned about a child's learning progress. The model then becomes very similar to what has been called "individualized" or "customized" instruction. This is a type of information which is likely to be highly useful to decision makers, and a major purpose of evaluation is to supply it. Seldom will information across more than one organization level be directly relevant to any decision process. Managers must deal at their own level of abstraction and operation; those who concern themselves with details down the line will lose their effectiveness.

The Stanford Evaluation Consortium (1976, p. 18) suggests that producing this type of information should be a prime concern in evaluation:

> Evaluation (in the new model) becomes a component of the evolving program itself, rather than disinterested monitoring. . . . Formal reports to outsiders are reduced in significance, and research findings become not conclusions but updatings of the system's picture of itself.

Information Which is Motivating.--To quote the Stanford group again, there is almost always a political aspect in programs and situations with which educational evaluation is concerned. Evaluators like to proclaim themselves producing data which will lead to rational decision making, and they are greatly disappointed when their outputs seem to be ignored. They are often tempted to the conclusion that decisions about education are irrational or, at best, non-

rational. This is a serious mistake, for all decisions are rational from the point of view of the person making them; it is, rather, a matter of which priorities are in operation. "Certainly, the evaluator who believes that the introduction of evaluation facts and findings will make an argument less political is dangerously deluded (Stanford, 1976, p. 12)."

What information does a decision maker use? Obviously, information is used which the decision maker perceives or has been conditioned to accept as important to his own concerns in the matter at hand. This is something like determining in behavioral analysis what will serve as a reinforcer. There is no doubt that human behavior is conditioned; what is a reinforcer for me may not be one for you, however, and specific reinforcers have different operational meaning for different persons. The same is true with different types of information.

The question at hand is what information will act as a motivator of human action, that is, be sufficient in Weaver's effectiveness or influence sense, and it seems quite clear that the answer is closely related to the notion of reinforcement. That is, the decision maker will be motivated by what he perceives or has been conditioned to accept as in his own best interests, as related directly to the goals he is seeking to achieve, as directly relevant to his own experiences.

This has to do immediately with the cross-levels concept, for the simple fact is that, in most cases, human beings will be less motivated by information removed from the level at which they are operating. To take an example with a time dimension, consider my concern for the different generations of my family. I take a great deal of interest in myself. I am also quite interested in my children and my parents. My interest lags a good deal when we shift to grandparents and grandchildren, and, the further we go in generations

from there, I have a difficult time getting very excited about it. I could care less, in fact, what may happen to the tenth generation down the line in 2177.

C. Suggestions for Action

To restate the hypothesis:

> The efficacy of information produced in an evaluation effort varies inversely with the number of organizational levels that the action the information describes is removed from the decision process the information is intended to influence.

To the extent that this is valid, evaluators would be well-advised to choose analysis units at the pertinent decision level or at immediately-adjacent levels. That is, they should produce information on the behavior or organizational units or systems at these levels. If you are in the US Office of Education, for example, concentrate on state-level actions and forget about trying to aggregate up from classrooms.

As corollaries to this major suggestion, we repeat two ideas from the conclusions of the Stanford Group (1976, p. 18).

- The evaluator should not concentrate on outcomes; ultimately, it may prove more profitable to study just what was delivered and how people interacted during the treatment process.
- The evaluator should recognize (and act upon the recognition) that systems are rarely influenced by reports received in the mail.

V. Recommendations

From workgroup reports and from consideration of the cross-levels hypothesis, sets of conference recommendations have been formulated, as follows:

A. Humanizing Evaluation

1. To the fullest extent possible, evaluation should be descriptive, positive and option-enhancing rather than judgmental, negative and option decreasing.

2. Evaluation should be significant and applicable to the needs of those being evaluated. If data from an evaluation program are not acted upon within a reasonable length of time, the evaluation should be terminated.

3. Evaluators should be concerned with disrupting the educative process and should collect the minimum amount of data compatible with producing appropriate and relevant information.

4. Research and evaluation should be conceived as separate concerns, and an attempt should be made to define the purposes and skills needed for each. In addition to technical skills, for example, skills in problem identification and solving, interpersonal interaction and political awareness are necessary for carrying out a successful evaluation.

5. Greater emphasis should be placed on gathering descriptive data about school settings prior to the evaluation of the effect of particular changes or interventions.

6a. Evaluators should use a variety of analytical techniques (e.g., anthropological and convergent techniques) and commit themselves to the complexity of their evaluative task, rather than relying on only one level or type of analysis.

6b. Evaluators and educators should take responsibility for updating evaluation procedures.

B. Involving the Evaluated

1. The evaluated community should have ultimate "ownership" of the evaluative process, and should have the right to challenge evaluation results it feels are invalid.

2a. Evaluators should combine a professional approach with a receptivity to pluralism. For example, evaluators should utilize their technical training to insure the validity of their instruments, while at the same time inviting the input and participation of parents, students, and teachers in planning and carrying out the evaluation. The Early Childhood Education (ECE) evaluation program is commended because it provides for local input and individualization.

2b. The role of the evaluator should be redefined so that he/she functions as an "ombudsperson;" i.e., a person who represents the public interest, is close to the public and is responsible to the public.

3a. Evaluators should strive for clarity to insure that all affected understand the meaning of evaluative discussions and reports.

3b. Evaluators should make their findings widely available, especially to people who might be able to learn from or apply these findings.

C. Racism and Sexism in Evaluation

1a. Evaluators should actively address themselves to reaching and evaluating the ways in which racism and sexism influence children's education.

1b. Evaluators should be aware of the diversity of ways in which individual children learn, and should develop evaluative and educative instruments which reflect this diversity. For example, evaluators and educators should challenge visual or linguistic materials which overrepresent males or confine females to stereotypical roles.

2a. Women in evaluation-related fields should caucus to insure that issues of racism and sexism are adequately addressed and that women's concerns are integrated into discussions of education and social policy.

2b. Women should be included in the dialogue at every level of evaluative decision-making. Being included means not only participating in the process of decision-making but also insuring that the terminology of decision-making is nonsexist.

D. Teachers and Evaluation

1. Teachers should not allow testing to determine education objectives or curriculum content.

2. Teachers should utilize evaluative instruments and techniques which deemphasize comparison. For example, criterion referencing can be substituted for IQ testing; also, teachers can evaluate class progress in terms of their own teaching goals and a child's progress toward understanding materials presented, rather than in terms of comparison between children.

E. Evaluation and Public Policy

1a. The humanistic evaluation constitutency should analyze and understand the problems faced by those evaluators who attempt to implement alternative evaluation techniques. Further, a concerted effort should be made to legitimize alternative techniques in the eyes of the evaluation profession, the public and state and federal funding agencies.

1b. Humanistic evaluators and educators should organize a public constituency for change of evaluation philosophy and practice.

2a. Evaluators should refuse to apply evaluative techniques which they believe to be ethically or professionally wrong, even if there is financial or

other pressure to apply these techniques. For example, the fact that federal funds may appear to depend on the application of inadequate evaluative instruments should not be considered sufficient ground to apply such instruments.

2b. Evaluators and educators should avoid being stampeded by bureauracy. For example, when a directive is received from federal, state, county or district level, evaluators and educators should carefully determine whether the directive is, in fact, mandated, or whether other options exist.

3. There should be concern with evaluating the quality of life in the schools, including such indicators as the number of children who have visual problems, who are hungry during the reading period and so forth. Further, a statement should be developed that would discuss the social outcomes attributable to this quality of lack of quality of life in the schools.

F. Increasing the Relevance to Decision Processes

1. Evaluators should choose analysis units that are at the same organizational level or at immediately-adjacent levels to that of the decision process the evaluation is intended to impact.

2. Evaluators should concentrate less on outcomes and more on what is delivered in a program and on how people interact as the program is implemented and operated.

G. Short Term Conference Follow-Up

1. Conference participants should make a personal commitment to return to their work places and attempt to do whatever thay can to make evaluation more useful and more humanistic. For example, they should stay in contact with each other to help solve common problems.

2a. Conference participants should try to organize local conferences in their own areas that could inform people about evaluation problems and alternatives and promote dialogue on the topic.

2c. A collection of documents or a bibliography outlining what is already known about evaluation should be distributed to all those interested, as should a resource list of evaluative models that represent alternatives to the use of standardized tests.

References

Miller, James G. "The Nature of Living Systems." Behavioral Science. 20:6 November 1975. pp. 343-365.

Riles, Wilson. Remarks at the 105th Charter Anniversary. University of California, Berkeley. March 29, 1973.

Shrode, William A. and Dan Voich, Jr. Organization and Management: Basic Systems Concepts. Homewood, Illinois: Richard D. Irwin, Inc., 1974.

Stake, Robert. "Evaluation Design, Instrumentation, Data Collection and Analysis of Data." Urbana, Illinois: Center for Instructional Research and Curriculum Evaluation, University of Illinois. (Mimeo.)

Stanford Evaluation Consortium. "Handbook of Evaluation Research." Educational Researcher. 5:10. November 1976. pp. 9-19.

Weaver, Warren. "Recent Contributions to the Mathematical Theory of Communications." In The Mathematical Theory of Communication. (Claude E. Shannon and Warren Weaver, eds.) Urbana, Illinois: University of Illinois Press, 1963.